

# Maskinlæring i NAV

*- hvilke problemer står vi overfor?*

//Cathrine Pihl Lyngstad & Robindra Prabhu,  
dataseksjonen NAV IT

Halvdagsseminar om bruk av maskinlæring i offentlig forvaltning,  
15.09.2021

1. Hvorfor og hvordan ML i NAV?



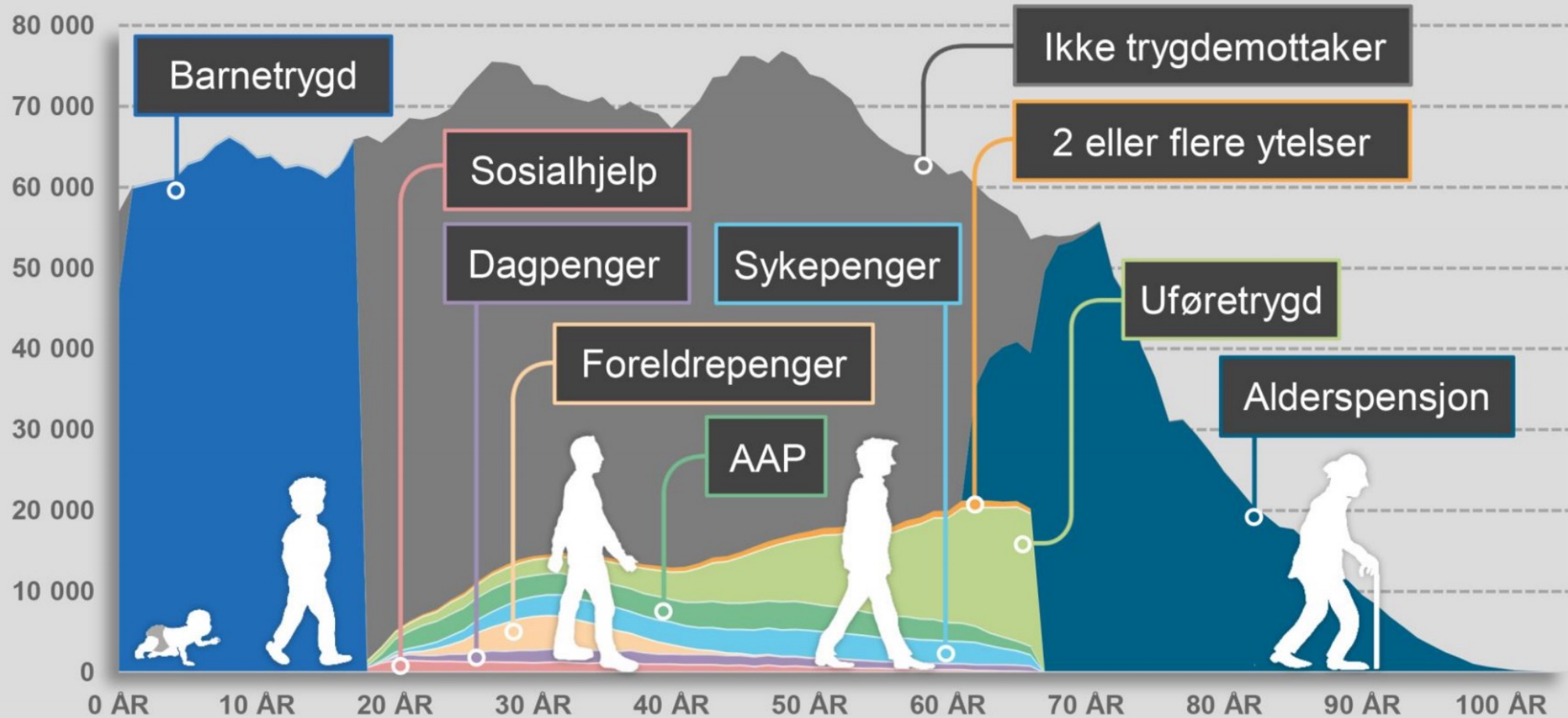
2. Tre nøtter til forvaltningsjuristen



# 1. Hvorfor og hvordan ML i NAV?



# NAV er med fra vugge til grav



# Hva skal vi med KI i offentlig sektor?




## 4.2 KI-drevet innovasjon i offentlig sektor

Regjeringen mener offentlig sektor har et stort potensial for å effektivisere og skape bedre tjenester gjennom digitalisering. Kunstig intelligens er en del av dette. Offentlig sektor vil i fremtiden bruke kunstig intelligens til å levere mer treffsikre og brukertilpassede tjenester, øke samfunnsnyttene av egen virksomhet, effektivisere drift og arbeidsprosesser og redusere risiko.


Kunstig intelligens i offentlig sektor kan blant annet bidra til:

- mer relevante råd og tjenester til innbyggerne i ulike livssituasjoner
- bedre beslutningsstøtte til saksbehandlere og andre offentlige ansatte
- å effektivisere prosesser og optimalisere ressursbruk
- å øke kvaliteten i prosesser og tjenester gjennom å automatisk oppdage sannsynlige avvik
- å forutsi trender basert på data fra både virksomheten og omgivelsene
- behandling av naturlig språk for sortering og kategorisering, og til å oversette mellom ulike språk og målformer


## NAV har en rekke aktuelle use cases for maskinlæring




Beslutningsstøtte ved oppfølging av arbeidssøkere




Tekstklassifisering for økt innsikt om brukerhenvendelser




Innsikt og prediksjoner knyttet til sykefravær



Målrette og prioritere tjenester mot arbeidsgivere



Matching av arbeidssøkere og arbeidsgivere/stillinger



Identifisere saker som bør behandles manuelt



# Case: beslutningsstøtte til NAV-veiledere



Vil den sykmeldte fortsatt være sykmeldt etter uke 26? ?

Ja, med 99% sannsynlighet

Utrekningen ble gjort Søndag 20. september 2020

Dette trekker varigheten opp	Dette trekker varigheten ned
↑ 1. Diagnose	↓ 1. Sykmeldingsgrad
2. Bosted	2. Lege
3. Yrke	3. Alder

[Detaljert informasjon](#) ▾

### Om faktorene

#### Diagnose

- hoveddiagnose (icpc og icd)
- symptom eller diagnose ved uke 17
- hoveddiagnosen med lengst varighet i personens tidligere sykefravær

#### Bosted

- kommunenummer
- gjennomsnittlig lengde på sykefravær for innbyggerne i kommunen
- arbeidsledighet i kommunen måneden før personen har vært sykmeldt i 17 uker

#### Yrke

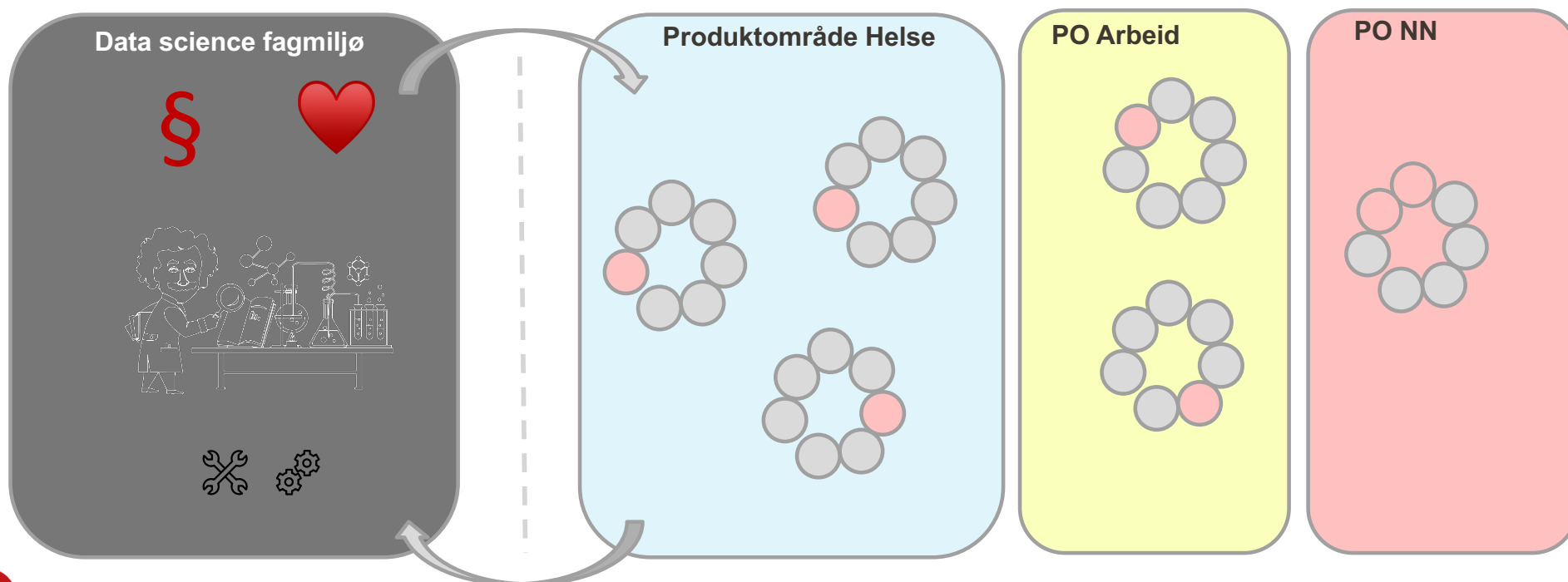
- personens yrke
- andre registrerte yrker
- gjennomsnittlig lengde på sykefraværet per yrke

#### Sykmeldingsgrad

- graden som brukes i sykmeldingen ved uke 17
- gjennomsnittlig sykmeldingsgrad fram til uke 17
- forholdet mellom sykmeldingsgraden i siste og nest siste sykmelding

# Matriseorganisering sikrer læring, utvikling og deling av god praksis

- Sentralt: Strategi, kompetanseutvikling og -deling, utvikling og deling av gjenbrukbare verktøy og metodikk
- Desentralt: Utvikling og bruk av maskinlæringsalgoritmer i tverrfaglige team



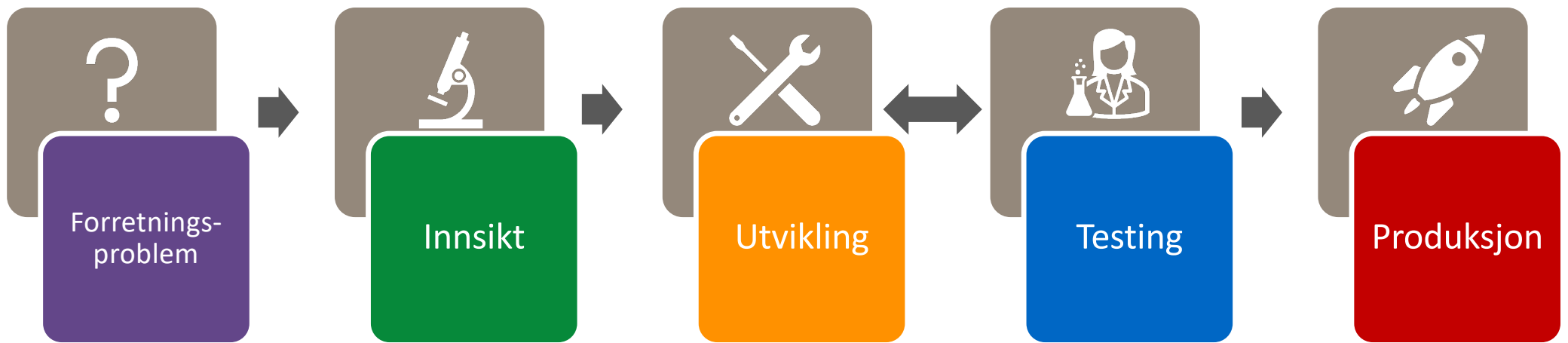


# Maskinlæring i NAV

*Hvordan jobber vi for ansvarlig utvikling og bruk?*

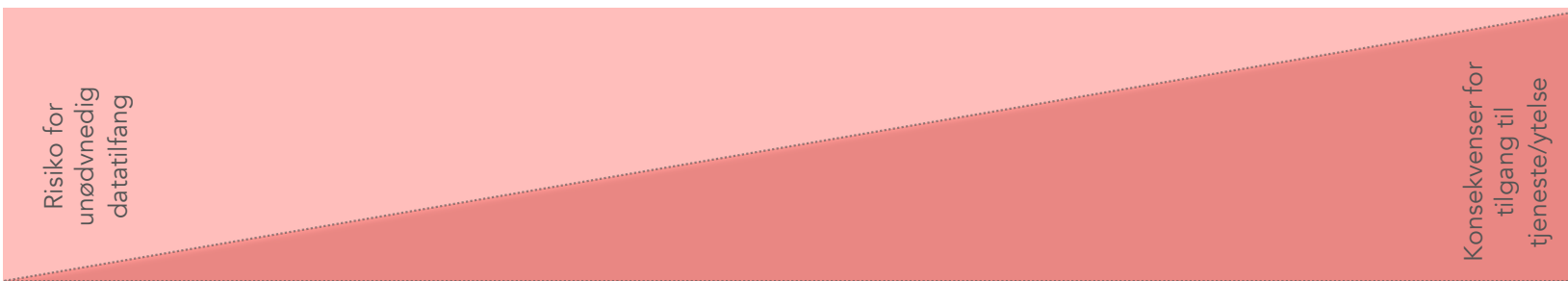
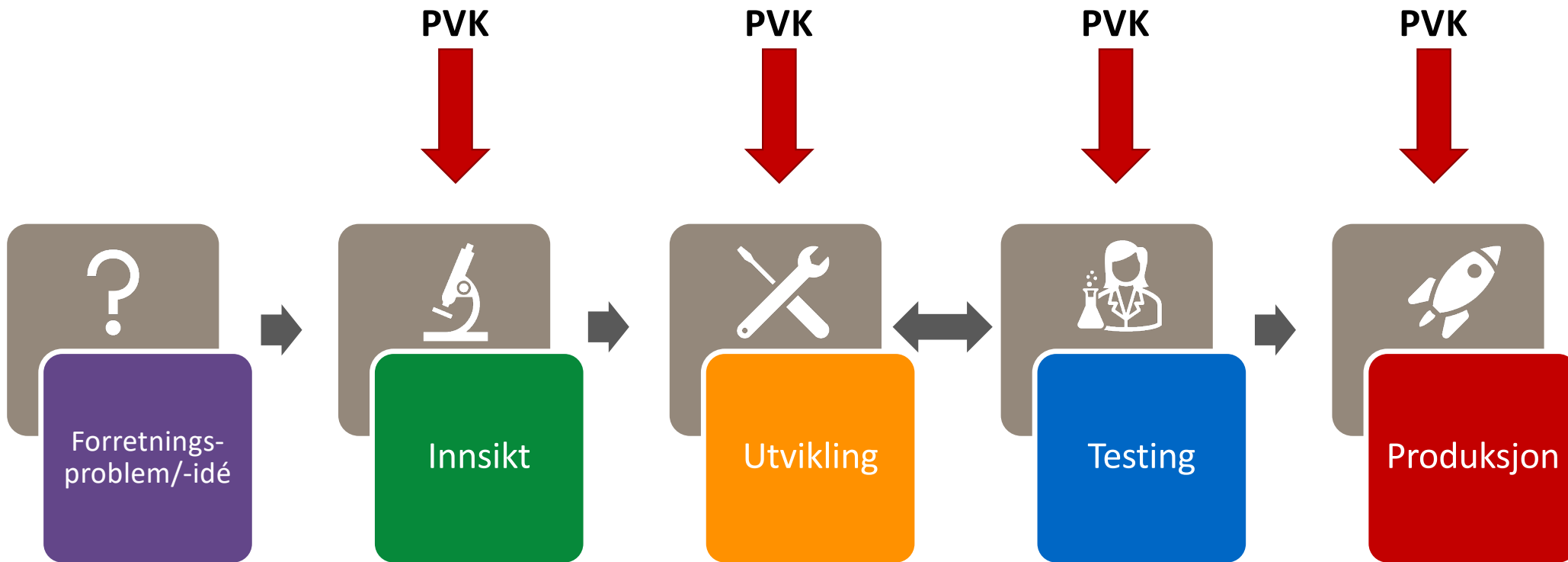
# Ansvarlighet handler om tankesett, støttet av gode prinsipper, verktøy og et tverrfaglig miljø for avklaringer og diskusjoner



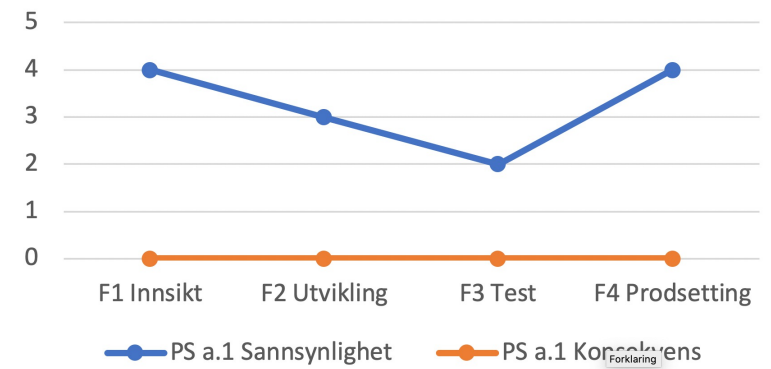
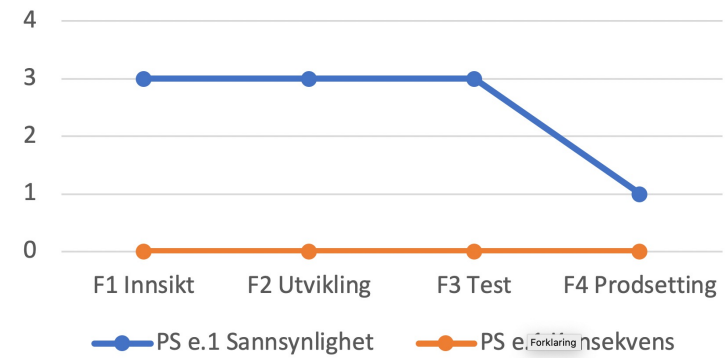
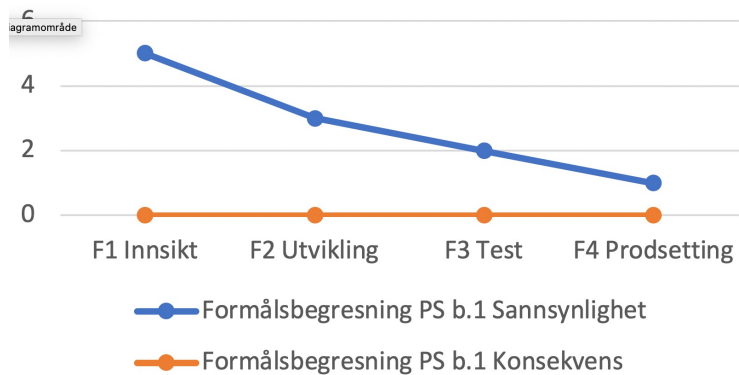
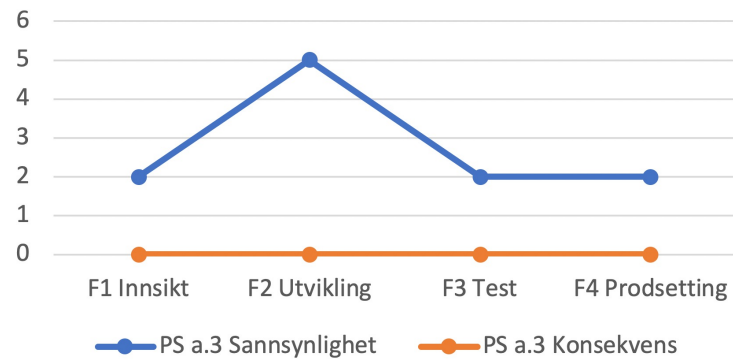


Wait! Has anyone  
done a DPIA?





# Personvernrisiko og -konsekvenser + evne til å vurdere disse endrer seg gjennom utviklingsprosessen



# Tre nøtter til forvaltningsjuristen



Har vi lov?





Har vi lov?

1

Hva er handlingsrommet for utvikling og bruk av maskinl ring i offentlig forvaltning?

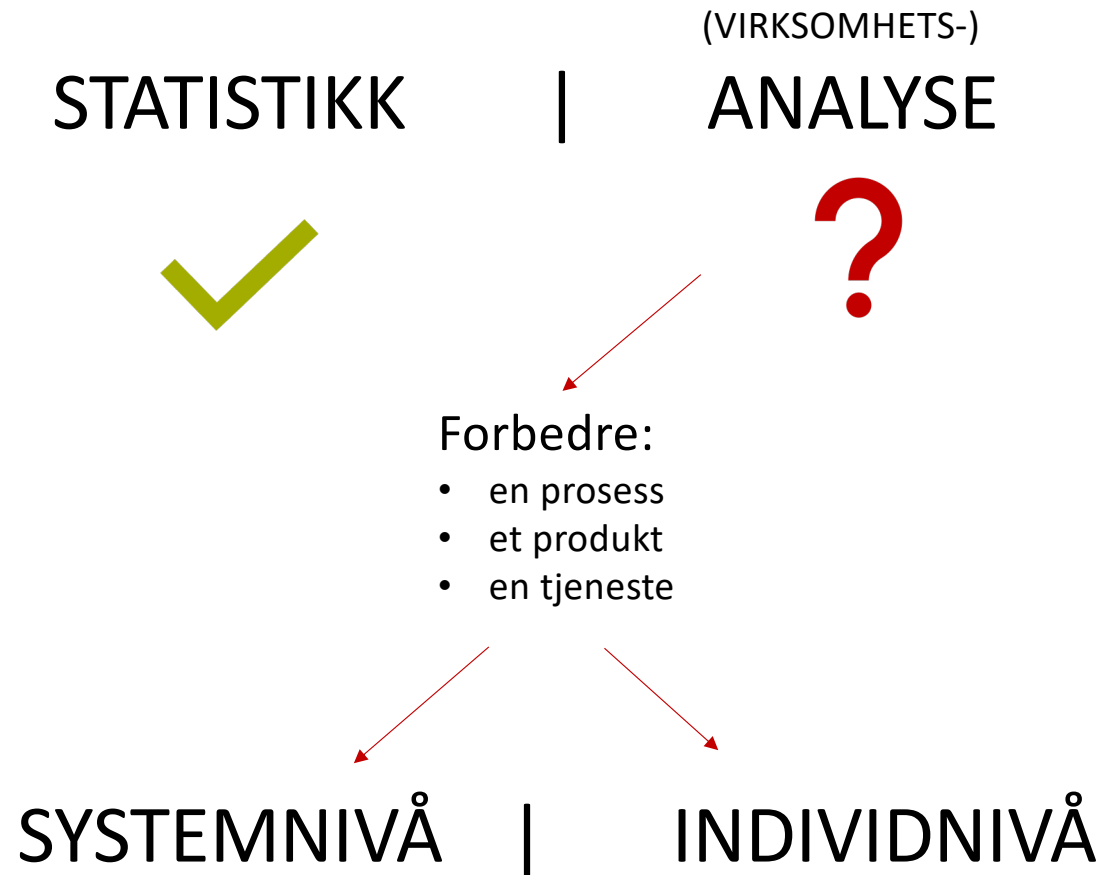
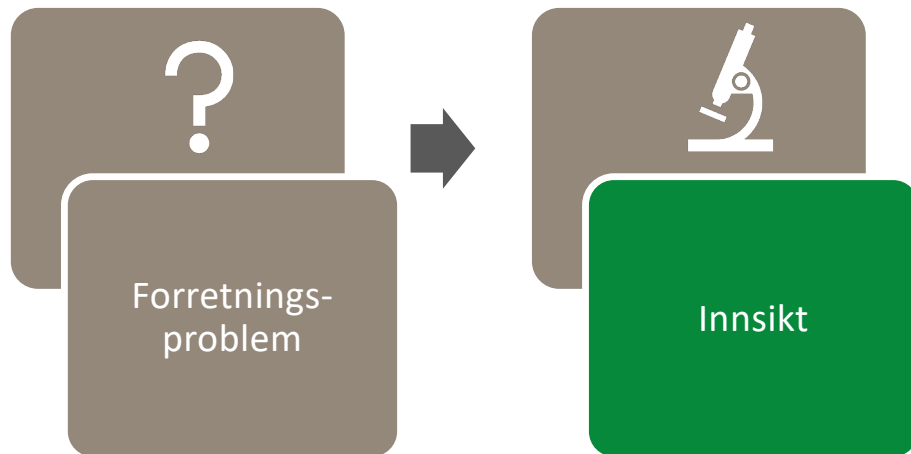
2

Hvordan b r utvikling og bruk av maskinl ring i offentlig forvaltning hjemles?



# Hvilket handlingsrom har vi for å ta i bruk KI?

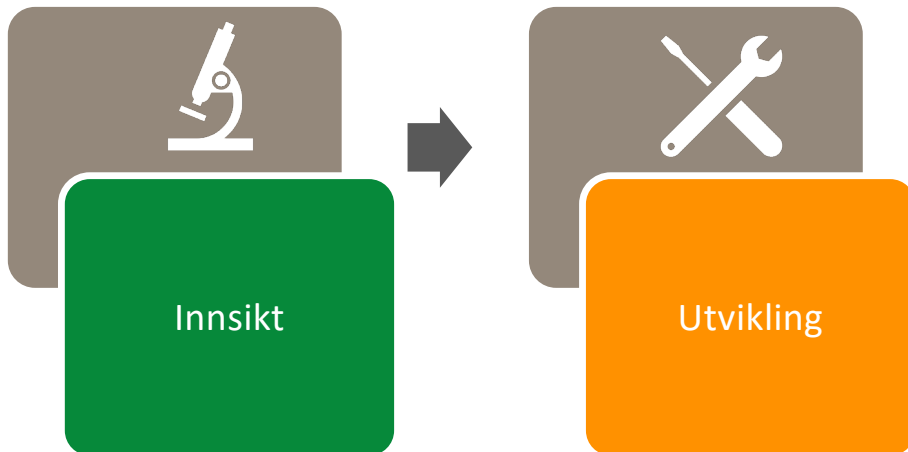
- prinsipielle avveininger i skillet mellom statistikk og analyse



# Hvordan bør utvikling og bruk hjemles?

- Når analyse glir over i produkt...

*I praksis: flytende overgang*



*I jussen: to vidt forskjellige ting?*

- Kan kanskje bruke data for retningsgivende analyse, men hva med produktutvikling?
- KI-produkter kan dessuten utfordre rettsikkerheten...
  - Taler for tydelige, forutsigbare rettslige skranker
  - Jo mer alvorlig / inngripende / «høy risiko» behandlingen er, jo tydeligere bør de rettslige rammene være

# Hvordan bør utvikling og bruk hjemles?

Jo mer *alvorlig* behandlingen er, jo tydeligere bør de rettslige rammene være

1

## Når er behandlingen så inngripende at særhjemmel påkreves?

*...sammenlignet med alternative løsninger vil KI nesten alltid innebære:*

- behandling av større volum PO
- behandling av PO fra andre registrerte som ikke er direkte saksrelevante
- (pseudonymisert) profilering

*Er maskinlæring alltid «for inngripende»?*

2

## Hvordan bør utvikling og bruk av maskinlæring i offentlig forvaltning hjemles?

- Forvaltningslov eller særlovgivning?
- På tjenestenivå («sykefraværsoppfølging») eller generelt (ala NAV-loven § 4a)
- Hva betyr «teknologinøytralt» i denne sammenheng?

## Tre nøtter til forvaltningsjuristen



Har vi lov?

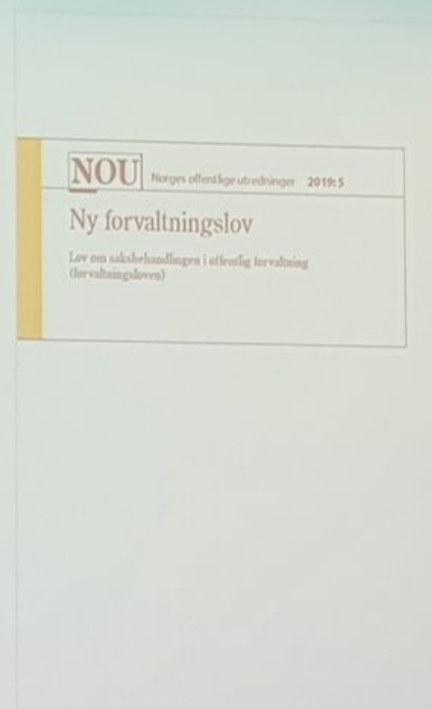


Rettferdig, og ikke-diskriminerende.  
Men ***hvordan?***



# Faren for diskriminering

Et automatisert saksbehandlingssystem vil behandle alle saker slik programkodene tilsier, og like situasjoner vil automatisk bli behandlet likt. Det kan imidlertid oppstå inhabilitetsspørsmål på utviklingsstadiet, selv om inhabilitet oppstår sjeldnere i arbeid med generelle saker. Et spørsmål som ikke gjelder inhabilitet, men som er beslektet, gjelder systemers egne skjevvrurderinger som følge av svakheter ved systemutviklingen eller mønstre i underlagsmaterialet som er tilgjengelig for systemet. Det er behov for å være oppmerksom på at også algoritmer kan være forutinntatte, diskriminerende og bidra til å opprettholde stereotyper.



**Datagrunnlag**

DATA GENERATION

POPULATION



Bias vil være en naturlig del av ML

- må håndteres i utviklingsprosessen for å unngå uønskede konsekvenser

**Historical bias**  
(normativ)

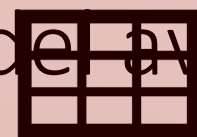
**Representation bias**  
(teknisk)

**Modellbygging**

DATASET &  
FEATURES

MODEL

MODEL

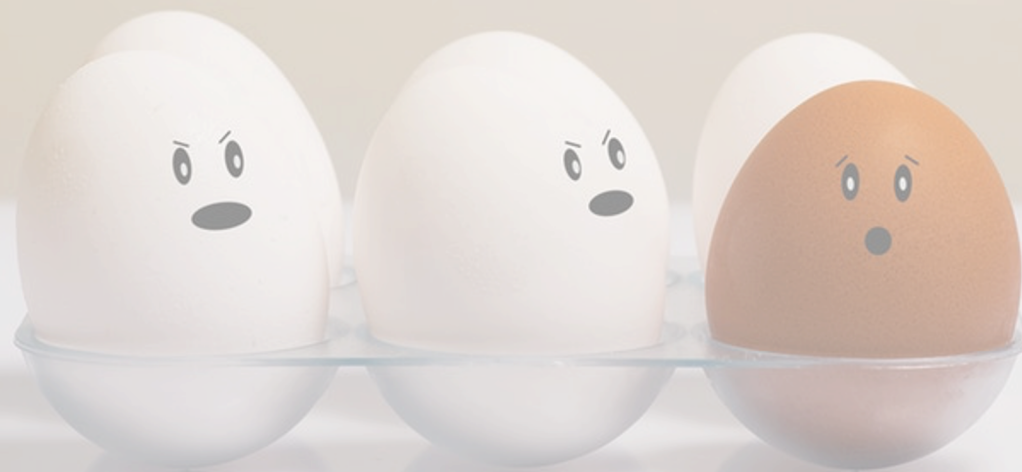


ML

**Measurement bias**  
(teknisk/normativ)

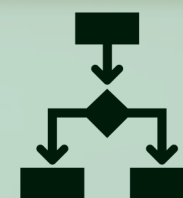
**Aggregation bias**  
(teknisk/normativ)

**Evaluation bias**  
(teknisk/normativ)



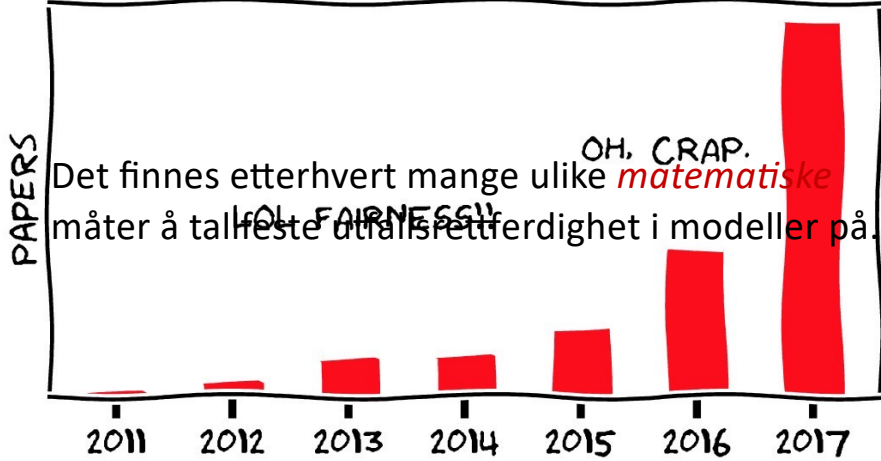
**Bruk**

DEPLOYMENT



**Deployment bias**  
(normativ)

## BRIEF HISTORY OF FAIRNESS IN ML



Credit: Moritz Hardt

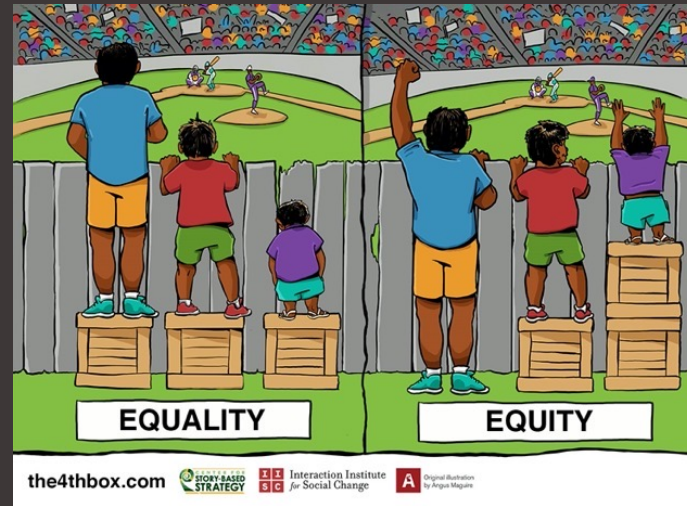
Demographic parity    Equalized odds    + +  
 Equality of opportunity    Conditional independence

Men...

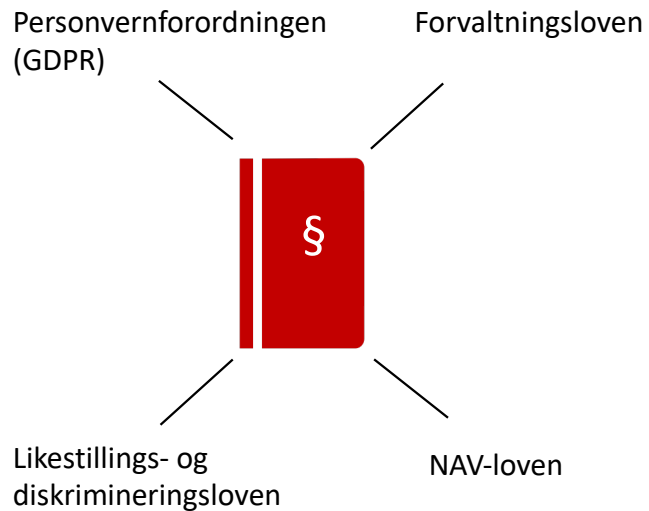
...disse kodifiserer ulike normative oppfatninger av rettferdighet

...valg av rettferdighetsmetrikk er *ikke* først og fremst en teknisk oppgave, men en komplisert moralteoretisk og juridisk øvelse som må gjøres i kontekst.

Og hva sier loven?



# Hva sier lovverk om *rettferdighet* i forvaltningen?



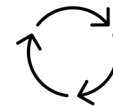
**Forskjellsbehandling**

**Diskriminering**

Individet



Prosess



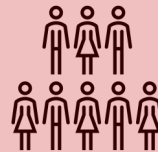
Saklig



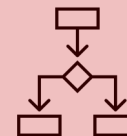
Direkte



Grupper



Utfall



Usaklig



Indirekte



FairML har hatt fokus (nesten) kun på dette!



# Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law

*West Virginia Law Review, Forthcoming*

51 Pages • Posted: 3 Mar 2021

Sandra Wachter

University of Oxford - Oxford Internet Institute

Brent Mittelstadt

University of Oxford - Oxford Internet Institute

Chris Russell

Amazon Web Services, Inc.

Date Written: January 15, 2021

## Abstract

Western societies are marked by diverse and extensive biases and inequality that are unavoidably embedded in the data used to train machine learning. Algorithms trained on biased data will, without intervention,

**Ikke tilstrekkelig å anta at inngangsposisjonen er rettferdig. Lovens intensjon er ikke bare å sikre *formal* resultatlikhet, men også "*substantive*" mulighetslikhet.**

**Må velge metrikker som er *bias-transforming*, ikke metrikker som bevarer status quo.**

classification system is intended to bridge the gap between non-discrimination law and decisions around how to measure fairness in machine learning and AI in practice. Finally, we show that the legal need for justification in cases of indirect discrimination can impose additional obligations on developers, deployers, and users that choose to use bias preserving fairness metrics when making decisions about individuals because they can give rise to prima facie discrimination. To achieve substantive equality in practice, and thus meet the aims of the law, we instead recommend using bias transforming metrics. To conclude, we provide concrete recommendations including a user-friendly checklist for choosing the most appropriate fairness metric for uses of machine learning and AI under EU non-discrimination law.

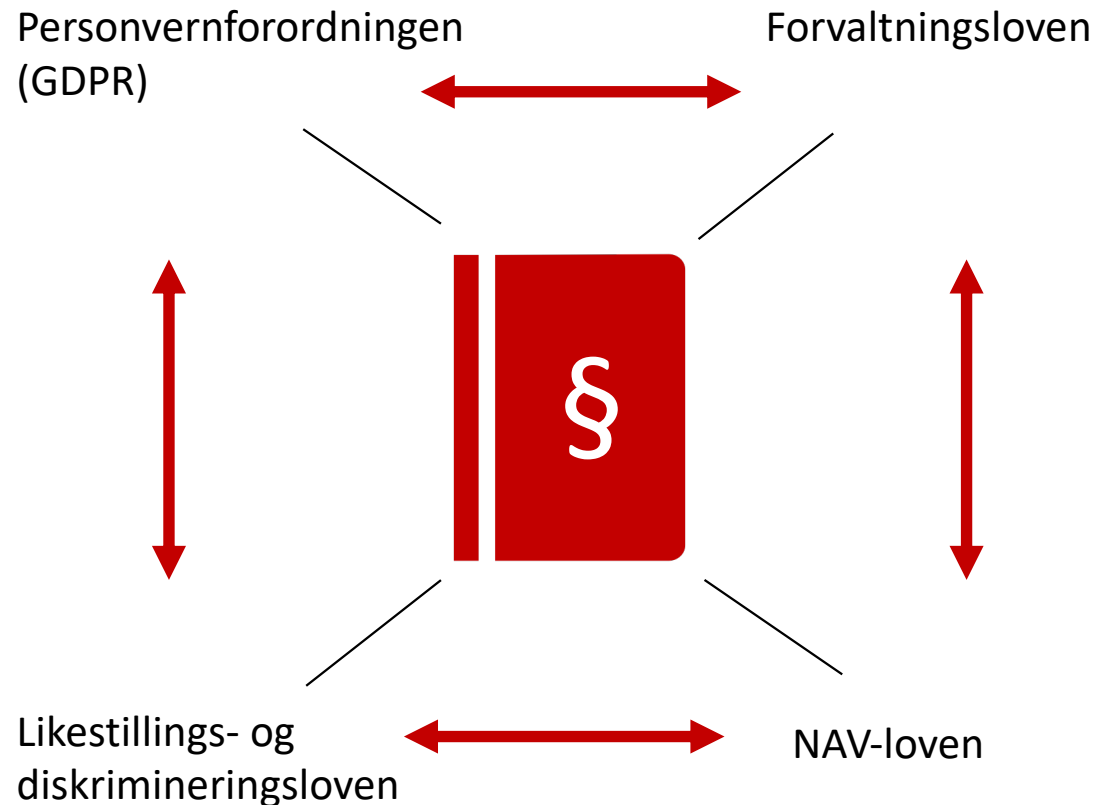
“Disse metrikkene harmonerer best med lovens intensjon”

Fairness metric	Bias preserving?
1. Group fairness, Statistical (demographic) parity	X
2. Conditional statistical (demographic) parity, Conditional independence	X
3. Predictive parity, outcome test	✓
4. False positive error rate balance	✓
5. False negative error rate balance, Equal opportunity	✓
6. Equalized odds	✓
7. Conditional use accuracy equality	✓
8. Overall accuracy equality	✓
9. Treatment equality	✓
10. Test-fairness or calibration	✓
11. Well-calibration	✓
12. Balance for positive class	✓
13. Balance for negative class	✓
14. Causal discrimination (direct discrimination)	*
15. Fairness through unawareness	*
16. Fairness through awareness	X
17. Counterfactual fairness	X
18. No unresolved discrimination	X
19. No proxy discrimination	X
20. Path based causal reasoning	X

**Table 1 – Bias preserving fairness metrics**

## Ok, men hva med andre relevante lovverk?

Kan vi oversette rettferdighetsbegrepet til prinsipielle føringer for design og utvikling?



## Tre nøtter til forvaltningsjuristen



Har vi lov?

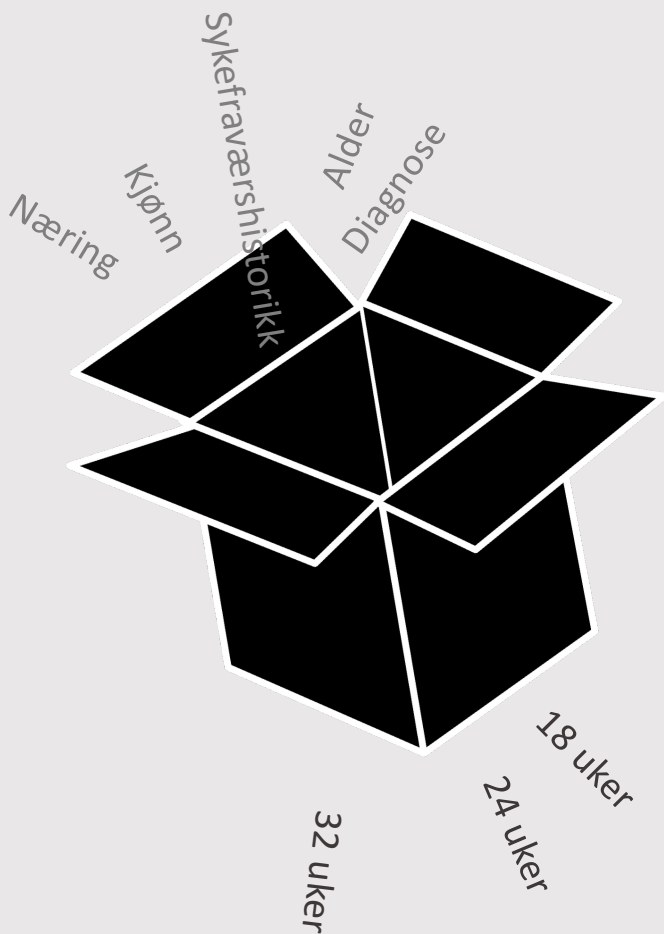


Rettferdighet og ikke-diskriminerende.  
Men *hvordan?*



Gjennomsiktig og etterretterlig  
Men *hvordan?*

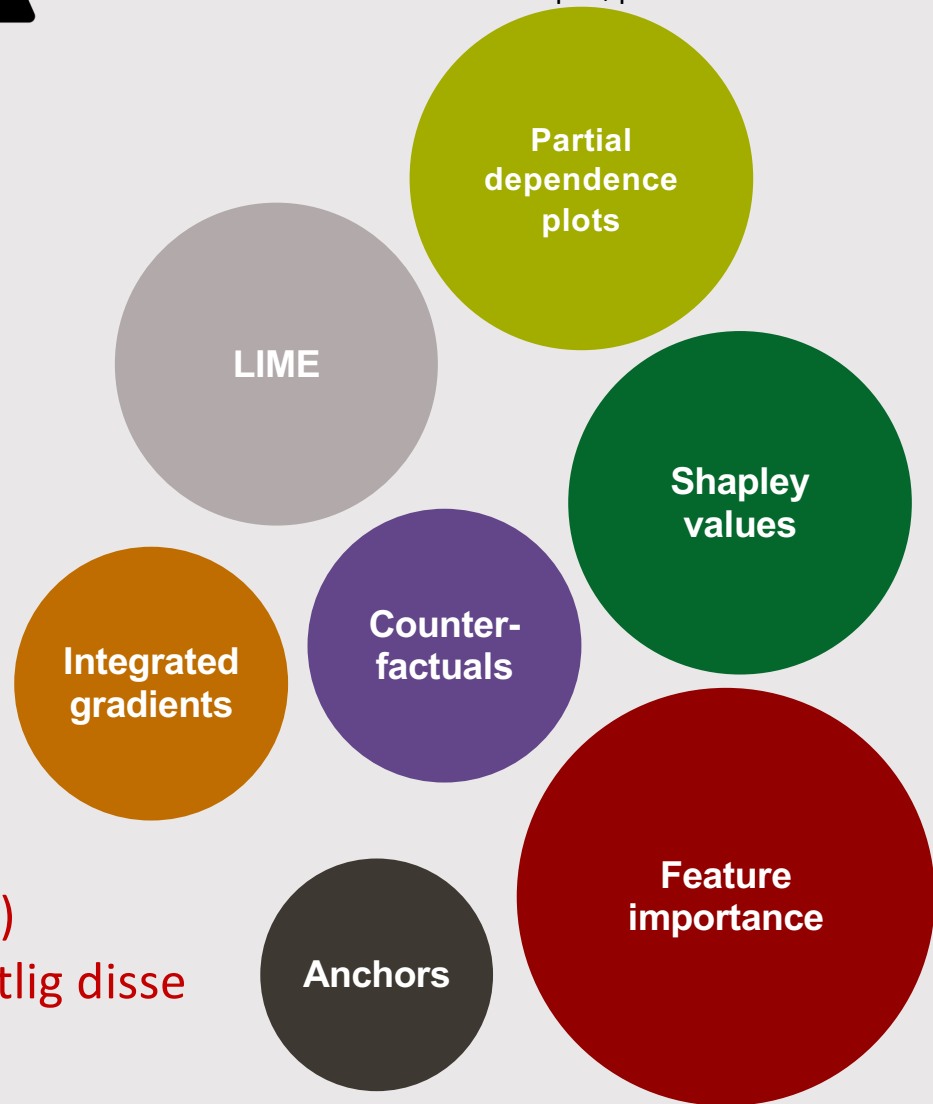
### 3c. Forklarbarhet



Men hvilket  
(forvaltningsrettslig)  
*problem* løser egentlig disse  
metodene?



Forskningen svarer på utfordringen  
- metoder for å forklare har kommet på løpende bånd



# Hva er transparens?

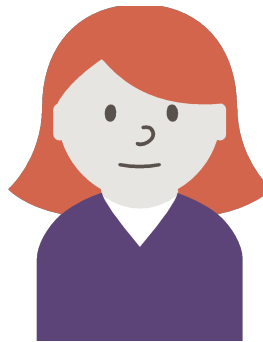
*Kommer an på hvem du spør!*

Ulike funksjonelle behov!



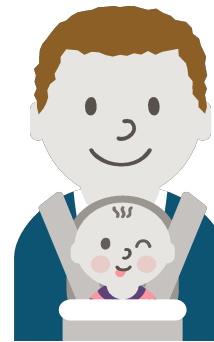
**Data scientist**

Forstå hvordan modellen oppfører seg, responderer på parameterendringer, hvor den gjør det bra og dårlig?



**Forretningssiden**

Passer den med forretningsmålene og er jeg rettslig compliant? Hva liker jeg til konkurrentene?



**Bruker / kunde**

Hvorfor ble søknaden avslått? Hvorfor fikk jeg en høyere pris enn naboen? Hva betyr det for meg? Hva kan jeg endre på?



**Tilsynsmyndighet**

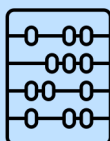
Tilgang til kode, kan jeg gjenskape resultatene, kan modellen gjøre til gjenstand for revisjon for å sjekke compliance?



# Forklarbar KI

- vi snakker om så mye forskjellig!

System



## Tolkbarhet (modellorientert)

*Kan vi forstå og forutsi hvordan modellen vil oppføre seg i ulike situasjoner?*



## Rettmessighet (justifiability)

*Er et modellutfall eller en modelloppførsel saklig og riktig?*

Individ



## Forklarbarhet (utfallsorientert)

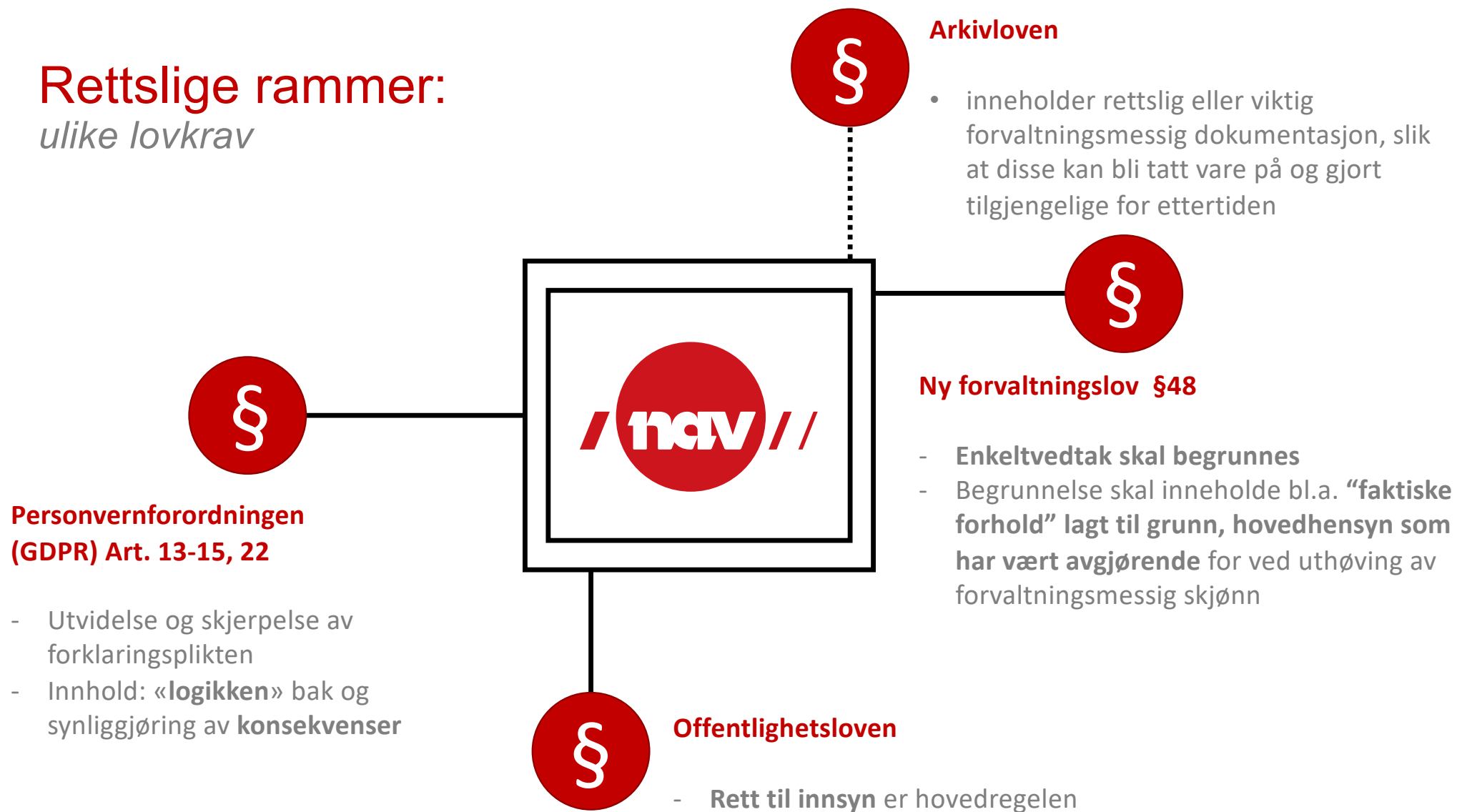
*Hvordan kom modellen frem til dette resultatet?*



## Forklaring (funksjonell)

*Har bruker nødvendig informasjon for å forstå, ivareta og utøve sine rettigheter / sin funksjon?*

# Rettslige rammer: *ulike lovkrav*



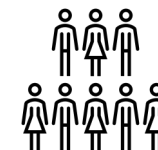


Så hva om vi spør  
juristen?

**PERSPEKTIV?**

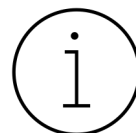


Enkeltsak

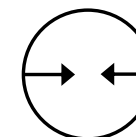


System

**RETTIGHET?**



Retten til  
informasjon



Retten til  
kontradiksjon

**NÅR?**

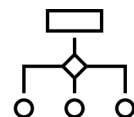


*ex ante*

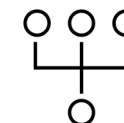


*ex post*

**HVA?**



Prosess



Utfall



UiO : Det juridiske fakultet

## Forklarbarhet i algoritmisk velferdsforvaltning

Demokrati og rettssikkerhet ved bruk av maskinlæringsmodeller i saksbehandlingen

U1

artikkel

## Prediksjoner av sykefraværslengde i NAV

En tverrfaglig case-studie om bruk av maskinlæring til å profilere

IOU

Jusstudentenes offentlige utredning

## Lex Digitalis

Hendelsesorientert tildeling av velferdsytelser

Hjelp oss å forklare riktig!

Loverk	Når?	XAI-metode
Offl § 3 allmennoffentlighet	Ex ante	global
Fvl § 16 forhåndsvarsel	Ex ante	global
Fvl § 17 uttalerett	Post hoc	lokal
Fvl § 18 partsinnsyn	Post hoc	lokal
Fvl § 11 veiledning	Ex ante og post hoc	global og lokal
Fvl § 23-25 begrunnelse	Post hoc	lokal

Skriv gjerne oppgave hos oss !



Rettslig grunnlag



Rettferdighet



Forklarbarhet

Takk for oss!